International Journal of Spatial Data Infrastructures Research, 2006, Vol. 1, 1-13.

[1]Introduction to the International Journal of Spatial Data Infrastructures Research

Max Craglia
European Commission – DG Joint Research Centre
Institute for Environment and Sustainability-Spatial Data Infrastructures Unit
TP262, Via Fermi 1, 21020 Ispra (VA), Italy
massimo.craglia@jrc.it

Abstract

This paper introduces the new International Journal of Spatial Data Infrastructures Research and set out its rationale and aims. The paper is organized into four main sections discussing the recent development of Spatial Data Infrastructures (SDI), the reasons justifying a specific research effort on SDIs, the areas of research that are currently central to their further development and use, and their relationships with research in the field of GI Science. An overview of the current research effort at the Joint Research Centre in this field is also presented.

Keywords: Spatial Data Infrastructures, research, social and economic impacts.

1. INTRODUCTION

There are several definitions of what constitutes a spatial data infrastructure (SDI) reflecting the variety of scope, level of jurisdiction, and thematic perspective. One definition that is often used is that of President Clinton's Executive Order 12906 establishing the (National) SDI in the United States, in which the NSDI is defined as "the technology, policies, standards, and human resources necessary to acquire, process, store, distribute, and improve utilization of geospatial data" (Executive Office of the President, 1994, Section 1-a). The importance of this particular definition is not so much in its content, which can be found with a high degree of similarity in many other initiatives, but in the degree of political support at the highest level of government that it embodies. The introduction to the Executive Order is also worth recalling as it captures many of the elements that have informed the discussions and activities related to SDI development since.

> *Geographic information is critical to promote economic development, improve our stewardship of natural resources, and protect the environment.*

> *Modern technology now permits improved acquisition, distribution, and utilization of geographic (or geospatial) data and mapping. The National Performance Review has recommended that the executive branch develop, in cooperation with State, local, and tribal governments, and the private sector, a coordinated National Spatial Data Infrastructure to support public and private sector applications of geospatial data in such areas as transportation, community development, agriculture, emergency response, environmental management, and information technology. (Executive Office of the President, 1994)*

The paragraph above shows a number of important elements, namely:

1) that GI and (Internet) technology have an important contribution to make to economic development and environmental policy,
2) that the political visibility of the NSDI arises through the National Performance Review which had been headed by Vice-President Al Gore with the title and mission: "From Red Tape to Results: Creating a Government that Works Better and Costs Less" (Gore 1993)**.** Hence, the earlier activities aimed at coordinating and developing a strategy for a SDI had found their political visibility and champion by becoming aligned, and be seen to contribute to, an important political agenda of the Clinton–Gore administration. From here the emphasis in subsequent sections of the Executive Order on the NSDI as a mechanism to "avoid wasteful duplication of effort and promote effective and economical management of resources".
3) that the NSDI needs partnerships across multiple levels of governments (and the private sector and academia) and strong coordination.

In practical terms SDIs include technical activities related to the dissemination and integration of key datasets, their documentation with metadata so that they can be found more easily and utilized better, standardization activities, and internet-based services to allow the discovery, access, and utilization of these key datasets. These central components have absorbed much of the implementation and research effort, but it is important not to loose sight of the dimension of SDIs relating to multiple partnerships, and their political nature as government-led activities that involve the sharing or redistribution of resources across many stakeholders. People, including education and training, and organizational capacity building, as well as coordination are other critical components of an SDI that are often overlooked (see also the editorial by Williamson et al. 2006).

Having set out some of the key features of SDIs, why should we research them, what are the main areas of SDI research, and how does SDI research differ from GISystems or Science research?

2. WHY SDI RESEARCH

A starting point to answer this question is that SDIs are a global phenomenon, with many initiatives not just in the developed countries but also in the developing ones [see for example Masser (1999, 2005), Williamson et al. (2003), Craglia et al. (2003), Vandenbroucke (2005), Crompvoets and Bregt (2003), Georgiadou et al. 2005)]. The sheer size and diversity of this phenomenon, which is increasingly involving not just national government agencies but also regional and local ones and multiple stakeholders from the public, private, and voluntary sectors, deserves attention. As Masser (2006) argues in his editorial this involves studying the diffusion of SDIs, their evolution, hierarchy, and the complexity of data sharing in practice within and among organizations. At the same time, after more than ten years of SDI developments, there is a need to stand back and critically evaluate whether the promises made by their proponents have been delivered. As Georgiadou (2006) points out in her editorial we are still in the "mythical" phase of SDI development in which true-believers preach on the multiple benefits of SDI and related technologies to "improve" decision-making, save money, empower citizens and so on. Such claims have yet to be adequately researched. In this respect, an international workshop on SDI cost-benefit and return on investment (the first of its kind, which is telling in its own right) held at the Joint Research Centre in Ispra in January 2006 found that the overall situation has not progressed much from the description made by David Rhind, former Chief Executive of the Ordnance Survey of Great Britain, in the year 2000:

> *We know very little about how much money and other resources are actually being expended on maintenance of the existing national Spatial Data Infrastructures, let alone on creation of enhanced versions of them or who is providing these resources. In broad terms, we do not know whether these resources are being applied wisely. It would seem helpful therefore to carry out some sound accounting of this expenditure: arguments for adding to it or for using it more effectively or efficiently are unconvincing if we do not know the present practice (Rhind, 2000, pg. 53)*

Our collective lack of knowledge of the costs (human, social, economic, and environmental) and the benefits of establishing, operating, maintaining, and updating an SDI is due not only to the paucity of studies in this field, but also to the following other dimensions:

1) the majority of the studies available to date refer to the costs and benefits of setting up and operating a Geographic Information System (GIS) in a

single (or more rarely multiple) organisation. Although both GIS and SDIs share common ground in respect to the nature of the information used (geographic), it is questionable whether the findings of studies relating to GIS are directly applicable also to SDIs. What seems to be a distinguishing feature of an SDI is its distributed nature and connectivity via networks that makes it more difficult to identify ex-ante the user communities of the infrastructure. As these user communities become more diffused and varied, so the difficulty of identifying and quantifying the benefits of the SDI increases;

2) many of the studies are opaque on the assumptions made and methods used to derive the costs and benefits, or more generically the impacts of the infrastructure under study, so that the estimates made vary considerably, and it is difficult to understand who to believe and why. Even more crucially, whatever the assumptions made to arrive at these figures, there seems to be no monitoring mechanisms put in place to validate the assumptions made over time, and therefore contribute to knowledge in this field;

3) we lack a real understanding not only of how much an SDI costs, but also of the proportion of this cost in relation to existing investments in geospatial information, technologies, and other related SDI components. This is crucial for a proper assessment of the additional investments required by an SDI.

There is clearly a rich research agenda here for those with an interest in social and economic analysis, but also one that requires a better understanding of the components of an SDI (data, metadata, catalogues, network services and so on), their functional relationship and relative contribution to the overall costs and benefits, the relationships between SDI and related technologies: GI systems, information infrastructures, and underpinning Information and Communication Technologies, as argued by Nedovic-Budić and Budhathoki (2006) in their editorial. More crucially, we need to move beyond the purely economic accounting, to research the social implications and impacts of SDIs and related mechanisms of e-governance, including issues of control, potential loss of privacy, and marginalization of those that do not have the social capital, skills and knowledge to use the infrastructure, and challenge those that do. The lack of research on these social impacts alone should already make a strong moral argument for undertaking research on SDIs.

3. WHAT IS SDI RESEARCH?

Given the complexity and variety of SDIs, a more appropriate way to address this question might be to look at those areas of research that seem to be more critical at the present time. For convenience, I will organize these areas under

three headings: Software Architectures, Data and Methods, and Policies and Impacts. The intention is not to be exhaustive, but of reporting the outcome of several discussions on this topic held with colleagues over the past year.

3.1 Software Architectures

The current generation of SDIs is by and large still an extension of a desk-top Geographic Information System (GIS), addressing the same technical users. The focus of an SDI is on finding and accessing distributed data, rather than only using the data one already has, but the assumption is that the data once found and accessed is then processed by a technical user locally on his/her GIS. Services have emerged that allow typically to:

- publish and search geodata and GI services (Web Catalogue Service)
- access geodata (Web Feature Service, Web Coverage Service, Web Gazetteer Service and the Geographic Markup Language)
- visualize geodata (Web Map Service)
- transform between different coordinate systems (Web Coordinate Transformation Service)

These services are a step towards making SDIs more usable by the non-technical public, but are still data-centric, as reflected in the SDI name itself. Where we need to move towards, is a Service Driven Infrastructure in which the purpose is not to find data but to receive information, i.e. the result of processing done somewhere else by humans or machines that gives the user, whether technically competent or not, the answer to a question. This goal requires inter alia:

- translating a request into a workflow of GI services;
- finding services that qualify as parts of this service chain;
- comparing between different competing services
- ordering the execution of the service chain;
- receiving the answer shipped together with sufficient assessment information;
- allowing feedback on the answer received;
- staying informed about any new information.

Any of these activities requires research and testing, and some of challenges of chaining services are reviewed by Klien et al. (2006) and Kuhn (2003, 2005). The latter for example has been arguing that users of geographic information should be able to refer thematic data to semantic reference systems, just as they refer geometric data to spatial reference systems. Hence software should support:

- referencing and grounding process

- projections to simpler semantic spaces
- semantic translation among different reference systems.

Developing such semantic reference systems is one of the grand challenges that will underpin the next generation of SDIs. Similarly, if machine had to compare between alternative services, much richer information needs to be encoded into each service, including “softer” issues such as reputability and trustworthiness, applicability and relevance of a service implementing a particular process to the data in hand, negotiating multiple Digital Rights Managements signatures encoding the terms and conditions of each data item along the service chain; all of this across multiple (natural) languages, and (national) cultures as well as professional ones. Related areas of research under this heading include:

- Software architectures for interoperable spatial-temporal modelling and geo-simulation: interoperability considerations so far mainly address geodata access and simple geodata processing. Spatial-temporal models and simulations that drive many application are still far from being interoperable and component based (e.g. to easily link components of a hydrological models into a meteorological model), moreover ways for integrating existing and upcoming GI services with geo-simulations need further research. This of course also links to issues in service chaining highlighted above.
- Software (Geo)Agents: Software agents differ from conventional software in that they are long-lived, semi-autonomous, proactive, and adaptive and there is a clear potential in the context of addressing complex geoprocessing (through workflow and chaining), Geo-Semantics and spatial decisions support need to link to distributed GRID processing and work related to web enabling in-situ sensor network delivering dynamic data.

3.2. Data and Methods

Under this heading there are many areas of research that will be familiar to GI scientist as they include issues of data quality, generalization, visualization, analysis and modeling and so on. A particular angle that comes from researching these issues in the context of SDIs stems from the distributed and heterogeneous nature of the data, which requires particular efforts in respect to:

- methods for data mining, fusion, and integrating of distributed and dynamic data sources, including sensors (see for example Soille, 2004),
- scale, error, quality issues and generalisation of data from distributed resources;

- data harmonization and schema mapping (see for example Friis-Christensen et al. 2005),
- data tagging, discovery, filtering, and fitness-for-purpose assessments methods.

To give a flavour of the kind of research needed in respect to the last point above, the current situation is one in which the vast majority of geospatial datasets is not documented at all, hence the difficulty of finding it. If it is documented it is increasingly done with international standards (ISO 19115) that represent the traditional librarians' approach to documentation and management, typically one in which the metadata records are physically separated from the item they describe (traditionally a book). Whilst this model was fine in the analogue world, it is debatable why it needs to applied also to digital information resources rather than more modern approaches that automatically encode the metadata, or at least a large portion of it, directly into the data itself as in the case for example of digital photographs (see for example Tesic, 2005, Bulterman, 2004). Moreover, current metadata captures only the data producer perspective, and not that of the users. So active feedback from users as well as more automated means of clustering user preferences and searching, mining association rules, and deploying the results for the benefits of users as well as producers that are now standard practice among on-line retailers (e.g. Amazon) have yet to make it to the geospatial world (see for example Pike and Gahegan, 2004).

3.3. Policies and Impacts

Some of the research challenges under this heading have already been discussed in the Introduction and do not need to be repeated here. The critical starting point however is to develop a theoretical model informing the expected impacts of SDIs in economic and social terms, starting maybe from existing models developed for e-government (Corsi et al. 2006), or looking at related research in the field of Information Infrastructures (Hanseth and Monteiro, 2004) or social informatics (Kling, 1999; Kling et al. 2005). From such model(s), then one could develop metrics to measure impacts and appropriate methodologies to do so. It would be possible for example, to hypothesize that regional (i.e. subnational) SDIs by making local data more accessible to citizens and business, create a competitive advantage for information–based industries and value-added resellers than regions that do not have an SDI. Another hypothesis could be that there are positive impacts on governance through increased public participation. To test such hypotheses one would need to look at comparable regions and study them in some depth, bearing in mind that there is also likely to be an inverse cause-effect, in that regions with high installed capacity for innovation might be more likely to have also an innovative government investing in SDIs, or have higher capacity for public participation. So considerable care

would need to be exercised in selecting case-studies that have also comparable institutional and economic contexts. Nevertheless, if such research is not carried out, we will continue to hear unsupported claims about the alleged benefits (or costs) of SDIs.

Having set out some of the areas of SDI research that appear to be more urgent and promising at the present time, how do these differ from GI Science research?

4. SDI RESEARCH AND GI SCIENCE

Goodchild (1992), and Goodchild et al. (1999) define the field of GI Science as a subset of Information Science i.e. the systematic study according to scientific principles of the nature and properties of (geographic) information. The justifications for carving a sub-field focusing on GI include the specificities of GI in respect to statistical analysis, the particular issues it poses in respect to institutional, legal, and public policy perspectives, and the needs of GI technologies for basic research. Specifically in the context of the Varenius project, Goodchild et al. (1999) argued that a multi-disciplinary research effort was needed that addressed three main arenas: individual-level geographic cognition, GI systems and technologies, and societal impacts. From this perspective, SDI research is clearly part of GI Science as it addresses these very same issues although with a particular emphasis on the distributed/network dimension the technologies, data, and institutional dimensions/societal impacts.

4.1. Why a new Journal

The majority of current research in GI Science is characterized by a strong technical orientation and positivist paradigm (Georgiadou and Blakemore, 2006). We would expect SDI research to have a strong technical dimension but also to pay much greater attention to the relationships between technology, society, and governance given the inter-organizational, and political contexts within which SDIs are deployed and by which they are shaped. Hence, we would also expect socio-technical, critical, and reflective perspectives to have a very important role to play in the development of this field complementing the traditional positivist scientific paradigm.

The launch of this new Journal on Spatial Data Infrastructures Research aims at filling a perceived gap in the scientific literature and create a dedicated space in which this emerging body of research can flourish. The gap-analysis was undertaken in 2005 under three main headings: theme, quality, and scope. In respect to theme, a review of over 100 journals publishing papers on geo-spatial information, recognized that very few devoted any attention to SDIs. A detailed content analysis (Georgiadou and Blakemore, 2006) of over 1600 articles in seven leading GI journals confirmed this, and also gave supporting evidence to

the other two gaps identified. In respect to quality, few of the articles on SDIs appear in peer-review journals (they are mostly in conference proceedings); in respect to scope, both the GI academic literature and the SDI grey literature emphasize the technical dimension with little or no consideration to issues of impacts, social and organisational/institutional dimensions, and critical perspectives. For this reason, a new journal is needed that focuses on SDIs, gives ample space to critical and social perspectives as well as technical ones, and guarantees quality through a full peer-review policy. Furthermore, the new journal must support the wide dissemination of quality science through an open access policy.

4.2. Why at the JRC

The Joint Research Centre (JRC) of the European Commission is a large user of GI to support its scientific and operational activities particularly in the fields of natural hazards, land management, and climate change. The JRC also acts as technical coordinator of INSPIRE (Infrastructure for Spatial Information in Europe) This initiative is jointly coordinated with the Direction General Environment and the statistical office of the Commission, EUROSTAT (http://inspire.jrc.it). The technical role of the JRC includes research and development to ensure that the implementing rules of INSPIRE are sufficiently mature, stable, and tested to allow for a system-independent infrastructure to be operational across multiple languages and professional and legal practices, and that the grounds for the next generation of SDIs are prepared based on the experience developed. The JRC undertakes this role in liaison with the European and international research community, and the international standardisation bodies.

A dedicated Unit for SDI research and development has been established at the JRC with some 30 scientific staff, undertaking work in the following main areas:

- Methodologies for the creation of pan-European datasets through automatic mosaicking, feature extraction, and morphological analysis;
- Software architectures for SDI,
- Voice-enabled GIS in a mobile environment in the context of forest fire fighting,
- Data harmonization and schema mapping,
- Cross-catalogue search and retrieval and development of the European geo-portal,
- User needs analysis as foundation for the development of an internal SDI,
- Socio-economic impact analysis of SDIs,
- Contribution to the standardization work of international organizations (CEN, ISO), and industrial consortia (OGC).

Of course the JRC cannot do this work alone, and it relies on a network of partners through competitive projects like ORCHESTRA (software architectures), and SAFIR (Speech Automatic Friendly Interface Research), research collaborations with academic institutions (e.g. Zaragoza and Muenster), strategic alliances on the future directions of SDI development (e.g. with USA and Canada), operational joint working with the European Environment Agency, European Space Agency, European Geological Survey, and European Union Satellite Centre in respect to work of the standardization and specification community, and an expanding network of other scientific collaborations. One of the main advantages of the JRC is that it can lead the scientific research work on SDI, and test it and implement it at the same time both internally through the development of the European Commission SDI, and at the European scale through INSPIRE. The launch of this new journal is part of the research strategy to become one of the leading centres of excellence in SDI research. The Journal will not be a vehicle to publish the JRC work, but to attract the best ideas and experiences in this field worldwide, and contribute to the development of knowledge.

5. CONCLUSIONS

The paper has sought to provide the context within which the JRC is launching this new Journal of Spatial Data Infrastructures Research. To do so, it has identified why we need a specific effort on SDI research, what is particular about this field, and how it relates to the broader field of GI Science. Clearly, this is a personal perspective on the field, and in a way the shape of SDI research will define itself based on the contributions of the increasing number of researchers addressing this topic, or topics germane to it from which there is much to learn, such as social informatics, internet studies, information infrastructures, e-governance, and so on. The boundaries of what constitutes SDI research must remain loose to allow for the multi-disciplinarity and cross-fertilization that are critical to develop this emerging field. We have created the container, now it is up to you to fill it with quality work!

ACKNOWLEDGEMENTS

A particular warm thanks to all the colleagues in the SDI Unit of the JRC for the stimulating discussions on SDI research, and in particular to Lars Bernard who has filled the gaps in my technical understanding of the key challenges.

REFERENCES

Bulterman D. (2004) Is it Time for a Moratorium on Metadata? *IEEE Multimedia*, October-December, 10-17.

Corsi M., Gumina A., and C. D'Ippoliti (2006). eGovernment Economics Project: *Economic Model Final Version*. eGovernment Unit, DG Information Society, European Commission. http://217.59.60.50/eGEP/Static/Contents/final/D.3.3_Economic_Model_Final_Version.pdf [accessed on 25/7/2006]

Craglia M. et al. (Eds.) (2003). *GI in the Wider Europe*. http://www.ec-gis.org/ginie/doc/ginie_book.pdf [accessed on 25/7/2006]

Crompvoets J. and A. Bregt (2003). World status of national spatial data clearinghouses. *Urisa Journal* 15:43-50.

Executive Office of the President (1994) *Coordinating Geographic Data Acquisition and Access: the National Spatial Data Infrastructure*, Executive Order 12906 Published in the April 13, 1994, edition of the Federal Register, 59(71): 17671-17674. http://130.11.52.153/publications/documents/geninfo/execord.html [accessed on 25/7/2006]

Friis-Christensen A., C. Jensen, J. Nytun, and D. Skogan. (2005) A conceptual schema language for the management of multiple representations of geographic entities, *Transactions in GIS,* 9(3):345–380.

Georgiadou, Y. (2006). SDI ontology and implications for research in the developing world. *International Journal of Spatial Data Infrastructures Research*, 1, 51-63.

Georgiadou Y. and M. Blakemore (2006). A journey through GIS Discourses. *URISA Journal* (Under review) [accessed 30/1/2006].

Georgiadou, Y., Puri, S. K. and S. Sahay (2005). Towards a potential research agenda to guide the implementation of Spatial Data Infrastructures—A case study from India. *International Journal of Geographical Information Science*, 19(10): 1113-1130.

Goodchild, M. F. (1992) Geographical information science. *International Journal of Geographical Information Systems*, 6: 31- 47.

Goodchild, M.F., Egenhofer M.J., Kemp, K. K., Mark, D. M., and E. Shepard (1999). Introduction to the Varenius Project. *International Journal of Geographical Information Science*, 13(8): 731-745.

Gore A. (1993) *From red tape to results: creating a government that works better and costs less*. Report of the National Performance Review, Sept. 7. http://govinfo.library.unt.edu/npr/library/nprrpt/annrpt/redtpe93/index.html [accessed 4/8/2006[

Hanseth, O. and E. Monteiro (2004). *Understanding information infrastructure*, manuscript available on http://heim.ifi.uio.no/~oleha/

Klien E., Lutz M., and W. Kuhn (2006) Ontology-based discovery of geographic information services—An application in disaster management *Computers, Environment and Urban Systems*. 30(1): 102-123.

Kling R., Rosenbaum H., and S. Sawyer (2005) *Understanding and Communicating Social Informatics A Framework for Studying and Teaching the Human Contexts of Information and Communication Technologies*. Information Today Inc.

Kling R. (1999). What is Social Informatics and Why does it Matter? *D-Lib Magazine*, 5(1) http://www.dlib.org/dlib/january99/kling/01kling.html [accessed 27/7/2006]

Kuhn W. (2005) Geospatial Semantics: Why, of What, and How? *Journal on Data Semantics*, Special Issue on Semantic-based Geographical Information Systems, Lecture Notes in Computer Science, 3534: 1-24.

Kuhn W. (2003) Semantic Reference Systems *International Journal of Geographic Information Science,* Guest Editorial. 17(5): 405-409.

Masser I. (2006) What's Special about SDI Related Research? *International Journal of Spatial Data Infrastructures Research*, 1: 14-23.

Masser I. (2005) *GIS Worlds: creating spatial data infrastructures*. Redlands: ESRI Press.

Masser I. (1999) All shapes and sizes: the first generation of national spatial data infrastructures. *International Journal of Geographical Information Science* 13: 67-84.

Nedović-Budić, Z. and N.R. Budhathoki (2006) Technological and Institutional Interdependences and SDI – The Bermuda Square? *International Journal of Spatial Data Infrastructures Research*, 1: 36-50.

Pike WA and M. Gahegan (2004) Visualizing concept relationships in a distributed knowledge sharing environment. *GIScience*, Adelphi, MD, October 2004.

Rhind, D. (2000). "Funding a spatial data infrastructure", in Groot, R and McLaughlin, J. (Eds.), *Geospatial data infrastructures*. Oxford: Oxford University Press, pp. 39-56.

Soille P. (2004) *Morphological Image Analysis: Principles and Applications* (2nd Ed.) Berlin: Springer-Verlag.

Tesic J. (2005) Metadata Practices for Consumer Photos, *IEEE Multimedia*, July-September, 86-92.

Vandenbroucke D. (2005). *Spatial Data Infrastructures in Europe: State of Play, Spring 2005*. K.U. Leuven. http://www.ec-gis.org/inspire/reports/stateofplay2005/rpact05v42.pdf [accessed 27/7/2006]

Williamson , I. P. , Rajabifard, A. and A. Binns (2006). Challenges and Issues for SDI Development. *International Journal of Spatial Data Infrastructures Research*, 1, 24-35.

Williamson, I.P., Rajabifard, A. and M-E.F. Feeney (Eds.) (2003) *Developing Spatial Data Infrastructures – From concept to reality*, London: Taylor & Francis.